


Article

Automatic Scaffolding Workface Assessment for Activity Analysis through Machine Learning

Wenzheng Ying ¹, Wenchu Shou ², Jun Wang ³, Weixiang Shi ¹, Yanhui Sun ¹ , Dazhi Ji ¹, Haoxuan Gai ⁴, Xiangyu Wang ^{5,6,*} and Mengcheng Chen ^{5,*}

- ¹ School of Design and Built Environment, Curtin University, Bentley 6102, Australia; wenzheng.ying@postgrad.curtin.edu.au (W.Y.); weixiang.shi@postgrad.curtin.edu.au (W.S.); yanhui.sun1@postgrad.curtin.edu.au (Y.S.); dazhi.ji@postgrad.curtin.edu.au (D.J.)
 - ² School of Built Environment, Western Sydney University, Sydney 2000, Australia; w.shou@westernsydney.edu.au
 - ³ School of Architecture and Built Environment, Deakin University, Melbourne 3220, Australia; jun.wang1@deakin.edu.au
 - ⁴ School of Electronics and Information Engineering, Tiangong University, Tianjin 300387, China; 1610920828@tiangong.edu.cn
 - ⁵ School of Civil Engineering and Architecture, East China Jiao Tong University, Nanchang 330012, China
 - ⁶ Australasian Joint Research Centre for Building Information Modelling, Curtin University, Perth 6845, Australia
- * Correspondence: Xiangyu.Wang@curtin.edu.au (X.W.); mcchen@ecjtu.edu.cn (M.C.)



Citation: Ying, W.; Shou, W.; Wang, J.; Shi, W.; Sun, Y.; Ji, D.; Gai, H.; Wang, X.; Chen, M. Automatic Scaffolding Workface Assessment for Activity Analysis through Machine Learning. *Appl. Sci.* **2021**, *11*, 4143. <https://doi.org/10.3390/app11094143>

Academic Editors: Inhan Kim and Jung In Kim

Received: 2 November 2020

Accepted: 27 April 2021

Published: 30 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Scaffolding serves as one construction trade with high importance. However, scaffolding suffers from low productivity and high cost in Australia. Activity Analysis is a continuous procedure of assessing and improving the amount of time that craft workers spend on one single construction trade, which is a functional method for monitoring onsite operation and analyzing conditions causing delays or productivity decline. Workface assessment is an initial step for activity analysis to manually record the time that workers spend on each activity category. This paper proposes a method of automatic scaffolding workface assessment using a 2D video camera to capture scaffolding activities and the model of key joints and skeleton extraction, as well as machine learning classifiers, were used for activity classification. Additionally, a case study was conducted and showed that the proposed method is a feasible and practical way for automatic scaffolding workface assessment.

Keywords: scaffolding; activity analysis; workface assessment; video camera; machine learning; skeleton extraction

1. Introduction

With the rising competition in the construction industry nowadays, construction corporations' profits have largely reduced, however, project owners' demands are increasing both in quality and efficiency aspects [1]. The construction's progress is an essential part of the construction technologies improvement [2–4]. Once project delay or rework occurs due to unsound construction management, it may result in profit decrease or even direct financial and credibility loss. In the oil and gas industry, there is a huge demand for scaffolding work during the construction and maintenance stage of a liquefied natural gas (LNG) plant. In order to keep the LNG plants running smoothly and safely, regular scaffolding erection and dismantlement are conducted for periodical inspections and part replacement [5,6]. Currently, scaffolding work is often designed as a preliminary step before the start of main construction tasks and also it has strong connections with other onsite construction trades [7]. However, scaffolding suffers from low productivity and high cost in Australia. One interview conducted with 56 construction contractors shows that scaffolding represents one of the 16 types of most wasteful components of indirect construction cost and its expense accounts for about 12–15% of the overall project cost [7,8].

Thus, timely and successful delivery of scaffolding work becomes a vital starting point of an entire construction project and it is also a crucial aspect for progress and cost management.

Project monitoring and controlling (PMC) is a management process that monitors and controls an as-built project by periodically following and inspecting the project's indices such as progress, cost, workload, productivity and so on [9–12]. PMC is regarded as a fundamental and key aspect in construction operations [13]. Continuous and effective monitoring and assessment on productivity, progress and quality are core tasks in PMC, which significantly impacts on the project's ultimate success. The PMC provides an opportunity to be aware of the ongoing as-built status and to identify as-built progress delays and assists in launching amendments [1,14]. Flawless PMC can ensure that projects are in accord with the budget and schedule [15]. However, the process of PMC has not been fully automated and there is a lack of widely accepted industry benchmarks for accuracy measurement [16]. In the last decade, there has been a huge demand for accurate and fast automated monitoring methods in the industry of Architectural Engineering Construction (AEC) [17–19].

Activity analysis is proved to be a feasible and practical approach for monitoring onsite operations and analyzing the conditions causing delays or productivity decline [20]. Activity Analysis is defined as a continuous procedure of assessing and improving the amount of time that craft workers spend on one specific construction trade. This amount of time is referred to as direct work time. Direct work activity is the activity that construction workers directly place physical effort towards. In 2010, the Construction Industry Institute (CII) proposed a detailed guideline for activity analysis. According to CII, a workplace assessment is described as the first and fundamental step for the execution of activity analysis [21]. Figure 1 shows an example of a workplace assessment, which indicates the distribution of activity categories where a period is spent. A workplace assessment aims to reflect on construction productivity before the release of cost and progress reports. It is a practical procedure of measuring the activity rates of construction workers throughout a long period of time, which relies on a professional onsite supervisor as an observer to determine and classify the activities executed by construction workers. However, the current approach to workplace assessment remains manual inspection and faces many challenges: first, manual data collection takes extra labor to observe, record and analyze the operation, which is not cost-effective due to high rising labor costs; second, close manual observation may result in an abnormal reflection of workers' performance caused by the Hawthorne Effect that the workers adjust their behaviors or productivity as a result of the awareness of being observed; third, the manual interpretation takes repetitive and random observations and requires a long period of time, and the results heavily rely on the supervisor's own experience [20,22].

ACTIVITY ANALYSIS TABLE						
Name of Observer	XXX					
Date	23 / 01 / 2018					
Project Name						
Trades Observed	SCAFFOLDING					
Observation Records	Round 1		Round 2		Round 3	
Time Range	9 AM	10 AM	11 AM	12 AM	2 PM	3 PM
Direct Work	+++		+++		+++	
Prep Work						
Tools and Equipment						
Travelling						
Waiting/Idling						
Comments						

Figure 1. An example of a workplace assessment.

Current workplace assessments require time-consuming and labor-intensive manual observation. To address these limitations, efforts including different machine learning algo-

rithms to automatize this process have been made by scholars in the past few years [23–25]. Cheng et al. utilized both location information and the worker's body posture for automated activity analysis of the task level. Their method combines the data obtained from UWB for location tracking and the body posture data from a wearable 3-axial accelerometer [26]. However, this approach merely used one single location and body posture as the training sample to infer each activity category. The activity recognition between two activities including intra-class variation and inter-class similarity at the same location would be challenging. Khosrowpour, Niebles and Golparvar-Fard proposed a method based on depth cameras for activity recognition and applied a bag of poses histogram and Hidden Markov Model algorithm to distinguish each activity [22]. However, depth cameras are sensitive to sunlight and a large body of construction activities are located outdoor. Thus, their method is feasible for interior construction operations but it may not be applicable to the external construction environment, such as scaffolding.

To automate the workplace assessment process, human activity recognition (HAR) performs a core and fundamental procedure in this research. Human activity recognition employs machines to comprehend and categorize a series of human activities from different data sources. Based on the type of data source, HAR can be divided into a sensor-based approach and a vision-based approach. The former exploits data from wearable sensors and other sensors. The sensor is attached to human limbs and other body parts or the surroundings close to the human activity to collect the target's activity continuously. Sensor-based HAR produces one-dimensional data, such as a time-series signal. Accelerometer, gyroscope and magnetometer are three common wearable sensors which detect and recognize the user's activities by analyzing the signal deviation when different activities are performed. Other sensors such as Radio Frequency Identification (RFID), Global Positioning System (GPS) and Ultra Wideband (UWB) also can be integrated into smartphones, wristbands, helmets or safety vests and provides a user's trajectory, absolute or relative locations to infer the user's activity. However, sensor-based HAR suffers from either high cost or sensitivity to the external environment [27–29]. For instance, a gyroscope is relatively expensive and needs to be attached to at least two key joints for HAR; an accelerometer is sensitive to temperature. In addition, most of these sensors need to be powered by a battery.

The latter, vision-based approach, collects human activities with the help of visual devices, such as by video surveillance or a camera, and generates multi-dimensional data including 2D images, 3D images or video frames. Although this approach releases participants from wearing attached sensors and the image processing technique develops, it heavily relies on visual data quality [30–32]. Elements, including image resolution and illumination conditions, to some extent affect the robustness of recognition. On the other hand, vision-based devices have been widely used in construction and other industries for safety management purposes and the cost of surveillance cameras is relatively low compared to other sensors.

This paper presents an approach that utilizes onsite video surveillance to collect scaffolders' activities in real time. With the help of machine learning and computer vision through semantic interpretation of real-time video sequences, this approach enables managers and inspectors to quickly obtain scaffolders' work status and conduct workplace assessment, which facilitates onsite scaffolding productivity and progress monitoring. Our approach is distinct from previous research since we select to use RGB video cameras that are relatively inexpensive (<USD 50) as well as being extensively available and have been widely utilized not only in HAR, but also in real time monitoring, accident analysis and other fields. In addition, RGB cameras are applicable to outdoor environments. Rather than inferring the activity category from the information of location or trajectory, we propose to extract 3D body skeletons from RGB video sequences and use machine learning algorithms for activity recognition from body skeleton sequences.

The literature review of research interests in construction workers' activities is stated in Section 2. The methodology in Section 3 illustrates the overall research procedures from

data collection, data preprocessing and activity definition to the human skeleton and key points extraction as well as the 3D point lifting and projection. To validate and test the feasibility of our proposed method, a case study was conducted in Section 4. Subsequently, Section 5 presents the conclusion and discussion on the research limitations and challenges as well as future work.

2. Literature Review

The integration of high-resolution cameras, boosting capability of data storage and the availability of high-speed telecommunication over the past decade, has revealed rich content on the onsite construction operation. Today, cameras and other electronic sensors have been broadly applied for contractors and owners to monitor ongoing construction activities [33]. With the help of algorithms in computer vision, scholars have placed interest in research areas in construction workers' activity mainly including ergonomics posture, safety management and productivity improvement [34,35]. This section reviews the state-of-the-art techniques for activity recognition and these application domains in the construction industry in which researchers have made great efforts.

Thanks to the rapid development of sensor technologies in recent years, automatic HAR has received consistent attention and has been widely explored in many domains. As introduced in the first section, the technologies for HAR can be generally divided into a sensor-based approach and a vision-based approach.

The sensor-based approach, based on sensor modality, can be classified into three types: body-worn sensors, environment sensors and object sensors. Body-worn sensors are often attached to the human body to detect the movement of the human body by continuously recording the signal data. Accelerometer, gyroscope and magnetometer are the three most frequently used body-worn sensors. Their applications focus on the activities of daily living (ADL) and sports. Environment sensors measure the medium that interacts between humans and the environment. For instance, pressure sensors, temperature sensors and sound sensors can measure their corresponding environmental parameters. Environment sensors can be utilized by monitoring hand movement and user's daily activities and they have been used in the field of the smart home. Object sensors are installed on objects which are close to human movements in order to infer human activity from the detection of object movement. For example, drinking activity can be detected by installing an accelerometer on a cup. A GPS or RFID module is attached to a worker's helmet to monitor the user's location to infer the worker's activity status [36–38].

Ryu et al. proposed to recognize worker's activities by using wristband Inertial Measurement Units (IMU) and conducted a case study of masonry work [39]. Bangaru, Wang and Aghazadeh assessed the reliability of wearable IMU and Electromyography (EMG) for HAR in construction [40]. Bangaru et al. established an Artificial Neural Networks (ANN) based model for the classification of scaffolder's work by using the data retrieved from IMG and EMG sensors [41]. However, these studies face technical challenges in that the models were trained with at-rest activities and, as a result, merely standardized and repeated movements could be effectively identified and recognized and transitional signals or actions were excluded or could not be correctly distinguished. Our approach takes an image sequence of each activity as the input data, which includes both typical actions and transitional actions to address this limitation.

According to the data type, the vision-based approach can be generally divided into RGB data and RGB-D data. RGB data refers to images consisting of red, green and blue color bands in the spectrum and RGB data can be collected through an ordinary camera or video surveillance. RGB-D data is produced by RGB-D cameras which can not only capture the original RGB data but also collect depth information. Dang et al. [37] pointed out that RGB data has the advantage that it is extensively available as well as affordable and contains rich content on the subjects. While compared to RGB data, RGB-D data provides depth information, which can enhance the performance of HAR. The disadvantages of RGB-D data include computation complexity as well as high costs [42]. Scholars have

explored and developed several models for both RGB data and RGB-D data and tested these models with public or benchmark datasets [36,37].

2.1. Ergonomics Posture

Attentions have been paid to ergonomics issues in construction manual work. Since manual work in construction usually requires workers to perform repetitive activities with heavy workloads, working in awkward postures can result in fatigue, injuries or severe accidents. Researchers have been exploring automatic methods of recognizing ergonomic posture and many assessment guidelines which focus on work-related musculoskeletal disorders (WMSDs). For example, the Rapid Upper Limb Assessment (RULA) and Ovako Working Posture Analyzing System (OWAS) are followed by most researchers for the definition and classification of manual activities and postures [43,44]. Ray and Teizer investigated a method of real-time posture analysis for ergonomics training by using a depth camera [45]. Yan et al. explored the potential of ergonomic posture recognition via a 2D ordinary camera [46]. Zhang et al. analyzed the joint angles from 3D skeletons generated by multi-stage convolutional neural networks (CNNs) in order to recognize the movements of body parts [47]. Wang et al. analyzed scaffolding posture and activities by following the OWAS and optimized the scaffolding activities [48]. Yan et al. collected and processed motion data to warn of the hazardous pattern on the head, neck and trunk with the help of wearable IMUs, a type of physical sensor attached to body limbs or trunk [49]. Seo et al. combined a depth sensor and own-designed system for biomechanical analysis [50].

2.2. Safety Management

Additionally, researchers have made efforts to solve the safety concern about worker's construction activities. Violation of operational rules about safety management might cause serious equipment damage or even a fatal personnel accident. Exploring an automatic system for identifying warnings against unsafe activities is a meaningful direction in that it relieves the burden of onsite supervision and quickly detects potential hazards. Yu et al. selected three unsafe behaviors: leaning on handrails, dumping from height, and ladder climbing, and then measured the skeleton angles generated from a depth sensor to identify unsafe behaviors in an experimental environment [51,52]. Alwasel et al. presented a framework to identify productive and safe poses of masons by implementing video cameras and IMUs attached on two control groups and a kind of machine learning algorithm to support a vector machine was employed as a classifier [53]. Han et al. focused on ladder climbing, a specific construction activity with considerable risk, and managed to map the body joints onto a 3D space by capturing and extracting motion data from a depth sensor system [54,55].

2.3. Productivity Improvement

A few studies aimed to explore construction workers' activity analysis for productivity improvement. Luo et al. integrated an RGB image stream, optical flow stream and gray image stream and trained them separately with CNNs to achieve workforce activity recognition [56,57]. Peddi et al. proposed a system for measuring construction worker's productivity through analyzing worker poses and classifying worker activity into three domains: effective, ineffective, and contributory work [58]. Khosrowpour et al. employed a depth camera for activity analysis and developed a bag of poses activity classifier and a Hidden Markov Model (HMM) for classification [22]. Calderon et al. synthesized pose sequences for the vision-based activity analysis of an excavator [59]. Apart from a vision-based approach, a mechanical approach that basically attaches different mechanical sensors to the human body to capture the signals of body and limb movement can also be regarded as a useful tool for motion capture [60]. Passive RFID and IMUs including gyroscopes and magnetometers were studied as practical tools [61]. Joshua and Varghese studied activity recognition in construction by using accelerometers [62]. Although magnetic sensors were

also adopted, they are susceptible to metal surroundings, which is not applicable for scaffolding work [63]. To increase accuracy, Alwasel et al. combined IMUs with video cameras to study productive masons [53]. The mechanical method is feasible under lab environments but it might face difficulties while being widely applied to construction sites due to the inconvenience for clothes wearing and washing caused by the attachment of sensors to the body.

To facilitate the procedure of workplace assessment, many studies have been devoted to approaches to automate it. The positioning technologies including Ultra Wideband (UWB) systems, Global Positioning Systems (GPS) and Radio Frequency Identification (RFID) only enable researchers to track the locations or positions of construction equipment and workers [64–67]. Besides, computer vision via video surveillance became another option for object detection and tracking [68–70]. These approaches remain at the stage where they infer the pattern of construction equipment and workers by tracking their positions in certain construction areas, which seems to be applicable to the equipment such as trucks or earth excavators, but not effective to onsite workers [71]. Without interpreting the construction activity that construction workers performed, retrieving useful data for workplace assessment is challenging.

3. Research Design

This section illustrates the research design and main procedures. Figure 2 displays the research framework of this paper. For the data collection, a 2D ordinary camera is adopted as our video capture device. To recognize scaffolding activities, it is essential to define these activities in advance. We follow the principle of activity analysis and divide scaffolding activities into three categories. In the step of feature extraction, key joints of the human body are used as the feature to be extracted from every video frame. To decrease the impact of intra-class variation and inter-class similarity, a model for 3D pose estimation is proposed subsequently. Then discriminative classifiers are trained with our annotated data to detect and recognize scaffolding activities. The case study and validation is described in the next section.

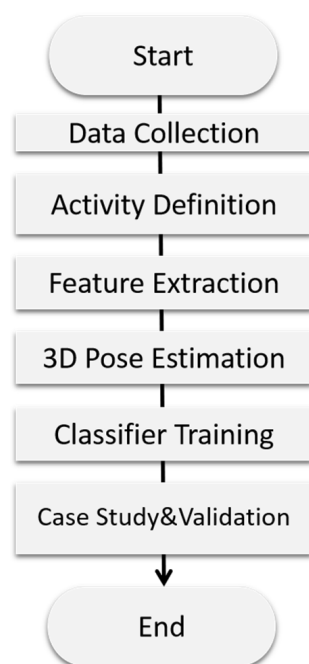


Figure 2. Research framework.

3.1. Using a 2D Ordinary Video Camera for Data Collection

Using an ordinary video camera for human motion capture has many advantages over the other data capture approaches mentioned above. Video surveillance employs ordinary video cameras and provides 2D videos to monitor and record activities and it has been broadly installed for management purposes in a large variety of industries. Compared to IMUs, an ordinary camera is non-intrusive: there is no requisite of devices attached to a worker's body; a depth sensor works properly under indoor conditions, but it is sensitive to the outdoor environment in that it produces numerous noises when it is exposed to outdoor radiation [47]. On the other hand, compared to depth sensors, an ordinary video camera is more cost-efficient and performs steadily on outdoor construction sites [47]. Furthermore, the video contains rich and intuitive information, which can not only be applied for monitoring, recording and safety management, but also can be broadly used for teaching purposes and accident analysis. To this end, the ordinary video camera was employed in this research as the human motion capture device for workplace assessment in construction [72].

Looking deep into human action estimation from video frames captured by a 2D ordinary camera, scholars have explored several principle approaches such as appearance-based [73], trajectories-based [74], volume-based [75], and interest-point-based [76]. As human body joint points contain rich, meaningful information for action estimation, the approach based on the extraction of key joint points has been widely studied in the field of computer vision and has become the state-of-the-art method for human activity recognition. Hence, the extraction of human key joint points and body skeleton estimation are adopted in our research for workplace assessment in construction.

This study proposes a vision-based method using 3D skeleton point estimation and supervised machine learning for workplace assessment from video sequences. For activity definition, we followed the guidelines for workplace assessments and divided the scaffolding operation into three categories: direct work, essential contributory work, and ineffective work. For model validation, we collected video data from actual scaffolding operations to train and test our model.

Intra-Class Variation

Ordinary cameras only provide a 2D video stream without depth information so that one major challenge in human activity recognition from 2D cameras is view invariance, which includes intra-class variation and inter-class similarity [47,77]. For intra-class variation, homogeneous activities might be identified as different ones since this class of activities are recorded and viewed from different directions. For inter-class similarity, heterogeneous activities may share similarities from certain viewpoints. For instance, in the Figure 3a,b inter-class similarity happens where the walking posture presents a similar skeleton appearance in 2D with the transporting posture. In Figure 3b,c intra-class variation takes place where the same activity of transporting is performed but filming from different viewpoints makes the variant skeletons. One robust human activity recognition method should have the capability to not only distinguish activities of different classes but also tolerate the intra-class variations in one homogeneous activity.

Compared with a 2D model, 3D human skeleton or joint positions can provide depth data from one more dimension to effectively decrease view variance and assist in the process of discrimination and classification. More detailed information can be provided by 3D features, which enables the process of action recognition to be more reliable and feasible. Thus, to enhance the accuracy and robustness, this study utilizes the extraction of a 3D human skeleton from 2D video frames captured by an ordinary camera for workplace assessment in construction.

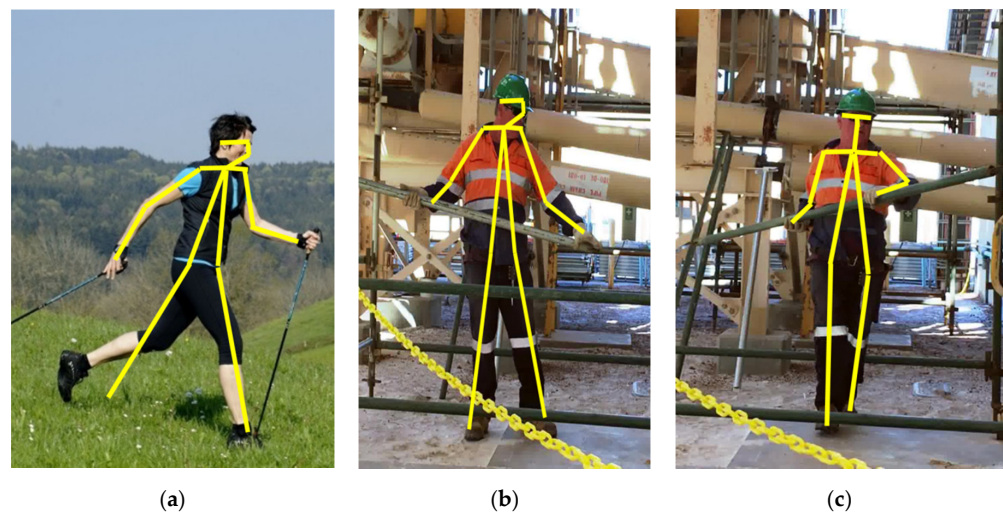


Figure 3. Inter-class similarity and intra-class variation: (a,b) represent walking and transporting, two different activities, while their skeleton appearance shows similarity from the same viewpoint; (b,c) display the same activity, while their skeleton appearances presents a variation from different viewpoints.

3.2. Activity Definition

Scaffolding operation is sophisticated and dynamic and scaffolding involves a series of activities and various body postures. However, it can be analyzed and simplified into a repetitive sequence of individual activities in accordance with the process of workplace assessment. For example, Khosrowpour et al. took interior drywall operation as a case study and divided the operations into seven categories [78]. In line with the principle of activity analysis, and with the assistance of interviews from onsite scaffolders and scaffolding supervisors, we analyze the scaffolding operation and categorize it into the three following sections below:

1. Direct work: the real process of contributing to a unit being constructed [20,79]. Additionally, from the thinking of lean construction, direct work is the process that adds value to construction work, which employers are willing to pay for [80]. For scaffolding operations, the part of scaffold erecting conforms to the feature of direct work;
2. Essential contributory work: the activities not directly set up but necessary to establish the construction unit. This category involves transporting materials and tools, receiving instructions, necessary communication between coworkers and so on. One typical essential contributory work for scaffolding work is scaffold transporting;
3. Ineffective work: the activities that contribute nothing to production probably due to inefficient material or labor supply and poor communication. These representative activities include idling or waiting;

From now on, we break down and investigate scaffolding operations with the aforementioned categories shown in Figure 4. and each category contains one representative: erecting (direct work), transporting (essential contributory work) and waiting or idling (ineffective work).

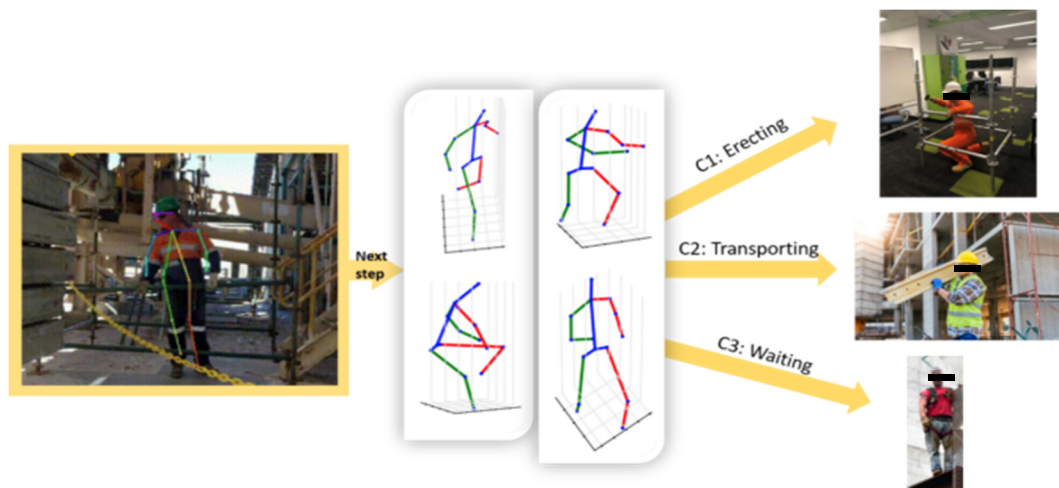


Figure 4. Scaffolding activity categories: Scaffolding activities are classified into erecting (direct work), transporting (essential contributory work) and waiting or idling (ineffective work).

3.3. Key Joints Extraction

This section illustrates the whole structure of 3D key joints extraction and posture estimation. In order to automatically recognize a scaffolder's posture under working conditions, we propose to integrate the OpenPose system and the 3D joint estimation model to extract 3D skeletons and key joints from 2D video frames. OpenPose is a multi-stage CNN system and represents state-of-the-art real-time model extracting human body key points from 2D video frames [81,82].

A human skeleton consisting of 18 joints is extracted by the implementation of the OpenPose system; a two-branch convolutional neural network (CNN), shown in Figure 5. Part confidence maps are produced by the first branch throughout Stage 1 and Stage t . Part Affinity Fields (PAFs) are generated by the second branch for limb association from Stage 1 and Stage t . Then the combination of part confidence maps and PAFs is parsed by the greedy algorithm to predict the 2D key points of the human body in the image.

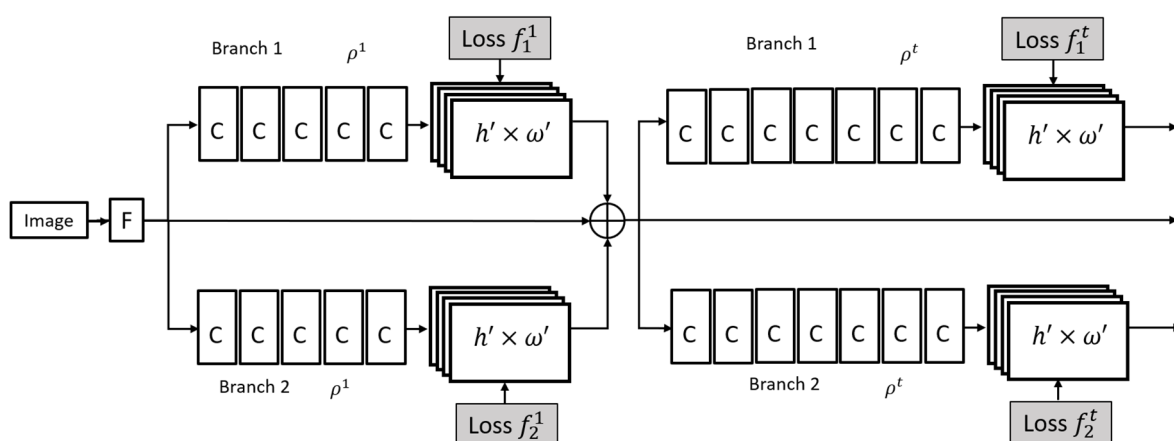


Figure 5. The structure of the OpenPose model.

The image is initially processed by a 10-layer convolutional network, where a set of feature maps F are formed [83,84]. As the input of stage 1, feature maps F are separately passed through two branches of CNN: branch 1 generates a set of parts confidence maps S_1 , which predicts human body parts in the image. Branch 2 forms a set of part affinity fields L_1 , which infer the association of body parts. Parts confidence maps S_1 and part affinity fields L_1 together with feature maps F are concatenated for the next stage afterwards. At

each following stage N , the prediction maps, including parts confidence maps S and part affinity fields L , are transmitted through two branches and concatenated with feature maps F for refining prediction. To iteratively train the multistage CNN, two loss functions are added at the end of each stage for each branch. The loss functions are designed to eliminate the difference between the prediction maps and the ground truth which is labeled manually and reflects correct body parts and part associations.

At every stage N , for each joint P this convolutional network generates belief maps for every pixel, indicating the confidence level that a joint point appears in any pixel (u,v) of one single image. At stage 1, the weights for each existing layer of convolutional networks are initialized by applying the weights of the Convolutional Pose Machine and those layers at the rest of stage N ($N > 1$) are randomly initialized. The architecture is trained through backpropagation by using the Human 3.6 M dataset, which contains 3.6 million human poses and corresponding 3D pose information [85,86]. For the conversion from pixel belief maps into body joint localization, the pixel with the most confidence level is chosen as the location of each joint.

3.4. 3D Pose Estimation

Under the feature extraction section, we first extract 2D key joints using the OpenPose model from every image frame and subsequently lift 2D key joints into 3D. Then the raw coordinates of the 3D joints are converted into relative coordinates by subtracting the x , y , z values of a central point under raw coordinates. Each joint point is separated into x , y , z three features and 42 features are used for machine learning classification. As shown in Figure 6, the 3D pose estimation stands on the basis of 2D human joint prediction. The belief maps produced by the OpenPose system are used as one of the inputs at each stage. In each stage the structure of 3D pose estimation merges (1) the belief maps from a 2D joint predictor (OpenPose system) and (2) projected belief maps generated by a 3D pose projection as the inputs. The process undergoes six stages. The 3D projection is designed to lift the 2D coordinates into 3D and project 2D point locations into 3D and then the fusion layer combines the 2D joint belief maps and 3D projected belief maps and propagates them into a set of 2D point landmarks in order to iteratively reinforce the 2D joint prediction as well as 3D projection. At Stage 6, the fused belief map becomes the final output and is projected to 3D models eventually. The architecture of 3D estimation is using backpropagation and trains the model end to end [47].

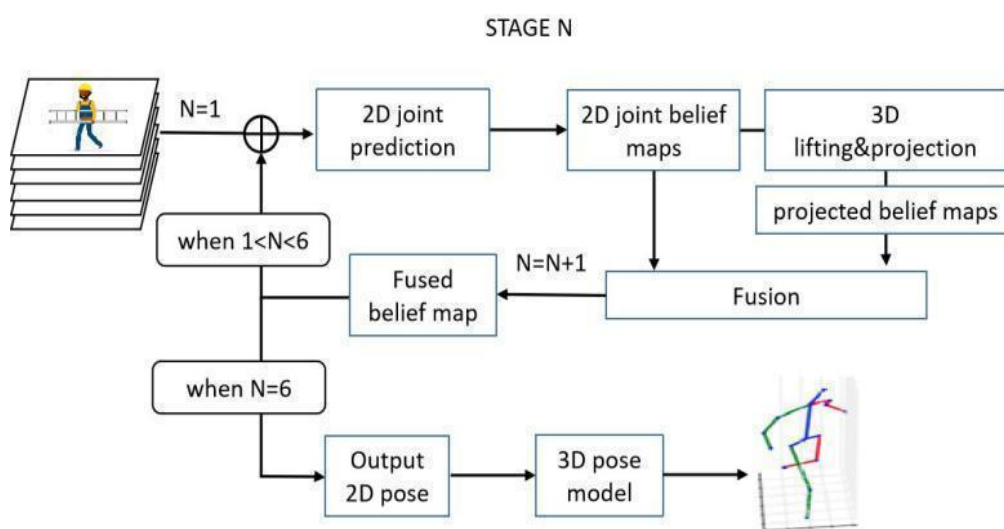


Figure 6. The structure of the 3D pose estimation model.

Inspired both by the approach [87] that represents the space of human poses as a mixture of Principal Components Analysis (PCA) and the idea that identifies poses as an

interpolation between pose categories, the probabilistic 3D model consists of a mixture of probabilistic PCA models with a number of clusters and the PCA models are trained by an Expectation-Maximization (EM) algorithm.

First, poses are subdivided into several pose categories C and the Euclidean distance d between pairs are computed. The objective is to look for a set of samples S that keeps the distance between joint points and their nearest sample minimized. This searching process is iterative and applies greedy selection that retains the previous S until another s additionally minimizes the Euclidean distance is found. The process stops when one's candidate becomes close enough to the existing candidate within little discrimination and the aligned points of a certain pose category are assigned to its closest candidate s and then the EM algorithm is executed and a mixture of probabilistic PCA bases is built accordingly.

$$\underset{s}{\operatorname{argmin}} \sum_{p \in P} \min_{s \in S} (s, p) \quad (1)$$

A unimodal Gaussian 3D pose model is implemented to estimate the 3D human pose of a single frame. This model in practice takes 20 samples and optimizes the 3D pose reconstruction for every single rotation to achieve rotational invariance on the ground. A non-linear least squares solver is applied to find the basis coefficients of the best-found solution and this approach provides results close to the global optima with the same average accuracy but less computational cost.

$$Y_p = \operatorname{argmax}_p b_p[u, v] \quad (2)$$

Then the generated 3D joint key points are projected onto a new set of 2D surface to form new 2D belief maps and, at the final layer in each stage, these new 2D belief maps, together with the belief maps from the 2D convolutional architecture, are fused into a single according to the equation below and the fused passed to the next stage as one of the input to refine the 2D joint prediction. For the final estimation of the pose, the 2D belief maps produced at Stage 6 are lifted into 3D space by the probabilistic 3D model [88].

$$f_t^p = w_t \times b_t^p + (1 - w_t) \times \hat{b}_t^p \quad (3)$$

3.5. Classifier Training

For the annotation, we collect a series of short video sequences with different durations for each activity category and each video sequence contains one single scaffolder conducting one type of activity. Additionally, based on our definition of scaffolding activities, we labeled these sets of video sequences with their activity categories: scaffold erecting (direct work), transporting (essential contributory work) and idling/waiting (ineffective work). The dataset details are presented in Table 1.

Table 1. Dataset details.

Activity Category	No. of Samples	Sum
Class 1: Working	811	1731
Class 2: Transporting	779	
Class 3: Idling	141	

As shown in Figure 7, we directly obtained 18 3D joint positions from the 3D pose estimation model. These joint points are under absolute 3D coordinates. Since the facial key points are not relevant and essential for the scaffolding activity recognition, for example, the locations of eyes and ears do not provide vital information for activity recognition, we simplify the points of eyes and ears by only keeping the nose location instead. This simplification can effectively increase accuracy and save the computational cost for classification. Since one single point in a 3D space provides x, y and z coordinate information, 14 3D joint points left from the elimination of 4 points of ears and eyes, comprise 42 features.

To facilitate the computation and make all the data comparable, we convert the absolute coordinates into relative coordinates by setting the central point as a zero point (0, 0, 0) and cutting off the same amount of x, y and z values, which are the initial values of the central point, from each key points respectively. We adopt the pose codebook that each pose of a kind of activity is viewed as a single histogram containing 42 features or vectors.

class	x0	y0	z0	x1	y1	z1	x2	y2	z2	x3	y3	z3	x4	y4	z4	x5	y5	z5
1	0	0	0	-123.009	4.039891	1.050438	-264.616	-442.641	-106.096	-209.266	-407.302	-481.461	109.4773	-4.03989	3.176284	282.6542	-461.301	-61.4646
1	0	0	0	-123.733	4.81664	0.672439	-266.419	-443.293	-106.836	-211.005	-405.939	-484.008	109.2881	-4.81664	2.789171	282.7132	-464.084	-62.1492
1	0	0	0	-119.748	-3.34965	2.62569	-251.269	-452.764	-102.388	-151.765	-410.841	-474.382	110.3013	3.349647	4.710845	287.3098	-447.941	-59.529
1	0	0	0	-150.361	-4.97629	2.787759	-242.773	-463.152	-97.827	-140.984	-429.72	-429.093	119.6792	4.976277	4.906296	300.6075	-459.002	-58.8866
1	0	0	0	-115.542	-3.82978	0.808794	-247.419	-458.373	-101.398	-143.142	-419.054	-433.332	117.5921	3.829774	5.733648	298.9911	-456.439	-59.2546
1	0	0	0	-110.395	-3.24758	0.922782	-236.169	-460.571	-52.8473	-131.878	-440.432	-418.264	122.2304	3.247569	5.733137	297.6316	-459.27	-53.589
1	0	0	0	-122.481	-13.0545	2.933637	-199.775	-478.53	-54.0747	-144.995	-465.02	-457.514	141.3068	13.05447	5.278184	327.3196	-446.354	-51.64
1	0	0	0	-115.294	-12.6677	-16.0707	-187.87	-467.486	-36.9349	-128.4	-444.292	-413.125	136.4526	12.66774	19.13157	308.4498	-440.943	-27.51
1	0	0	0	-132.526	14.11619	-18.0075	-281.719	-410.336	-86.3959	-167.965	-344.58	-444.835	131.898	-14.1162	27.05579	255.4249	-458.485	-122.2
1	0	0	0	-124.072	13.15002	-46.5411	-276.958	-388.647	-80.5397	-159.035	-314.068	-443.809	134.2473	-13.15	33.28558	262.1408	-440.387	-72.6374
1	0	0	0	-142.51	36.7136	-26.6462	-320.443	-310.202	-152.938	-194.423	-177.155	-482.085	118.8693	-36.7135	18.36021	235.4805	-437.537	-149.421
1	0	0	0	-135.413	29.50374	-25.523	-305.222	-336.71	-149.366	-180.683	-204.86	-475.568	128.0163	-29.5037	21.55687	248.4613	-441.076	-143.541
1	0	0	0	-93.5169	-87.6278	-4.95221	265.5818	-424.222	-120.833	291.8072	-373.374	-501.936	143.3239	87.62772	38.76596	354.9167	-299.239	-112.498
1	0	0	0	-99.9082	-59.7622	-73.4852	81.8245	-482.731	-74.9108	229.4437	-469.693	-426.138	155.5564	59.76211	7.782999	344.3944	-383.529	-144.148
1	0	0	0	-127.798	12.90975	-28.0482	-282.981	-404.249	-104.344	-165.502	-322.441	-471.641	135.8575	-12.9097	17.78221	265.5147	-453.196	-93.9869
1	0	0	0	-118.283	38.08803	-41.8376	-295.255	-316.994	-135.836	-213.854	-193.863	-518.637	108.1255	-38.0879	35.16317	221.8704	-441.267	-172.21
1	0	0	0	-134.965	20.20821	-25.4019	-254.045	-374.459	-149.82	-177.449	-252.165	-488.234	131.6358	-20.2081	17.57657	260.0097	-448.328	-146.298

Figure 7. Data frame of 3D joint positions. This figure partially displays the data frame of 3D joint positions. The actual data frame contains 43 features and more than 1700 lines.

To recognizing scaffolding activity from a short video sequence, a discriminative classifier is trained with the annotated data. The classifier takes the body skeleton features as the input, which provides sufficient descriptive content for activity classification. At the training stage, firstly we extract key points of a skeleton from every video sequence for each activity category and each video sequence is processed into frames which become the input of our pose estimation model. For every image frame, a set of 42 features generated by the pose estimator, plus one manually labeled feature indicating the activity category, constitute the input for supervised classifiers.

Since different supervised classifiers have various performances on one particular classification task, we introduce major popular multi-class algorithms including Random Forests (RF), Decision Tree (DT), Artificial Neural Networks (ANN), One-vs.-One Support Vector Machine (One-vs.-One SVM), One-vs.-All Support Vector Machine and K Nearest Neighbors (KNN) [89]. These algorithms are capable of multi-class classification as well as handling data with multiple features [89–91]. To select a classifier with outstanding performance, each aforementioned algorithm is employed for classification through our dataset of scaffolding activities and the training and testing process follows the principle of cross-validation which is a statistical procedure for the evaluation and validation of machine learning models.

Decision Tree (DT). Just like its name, DT is a kind of algorithm with a tree structure that a tree leaf denotes an outcome label, and a branch represents a sub-section of an entire tree or a sub-tree. DT can be used for classification and regression. For classification, a tree is constructed through the process of binary recursive partitioning, which is a procedure iteratively splitting the data into partitions. The Divide and Conquer algorithm is used in this process by breaking down a sophisticated group into two or more subsets with purely one type of feature, until the subsets become simple enough to be classified directly. DT is the building block of the random forest models [92].

Random Forest (RF). RF is an ensemble learning approach for classification and regression where, at the training stage, multiple individual decision trees are created by the selection of various subsets of training samples which follow the technique of bootstrap aggregating or bagging. The selection principle allows that the same data sample can be randomly selected several times while other samples may not be chosen at all [6,93]. At the prediction stage, every single tree structure independently produces a class prediction, and the ultimate prediction of the RF model belongs to the one with the majority of votes from decision trees. One single decision tree is susceptible to bias and variance, for instance, if one decision tree is too shallow its prediction is easily influenced by bias and if one tree is too deep, it is probably overfitting with a high variance. Since RF models aggregate a

multitude of decision trees, the models effectively reduce variance and mitigate volatility and noise because of data samples, which enhances the robustness of classification [94].

K Nearest Neighbors (KNN). KNN is intuitive, it is designed to search for the K closest data points to our target to be classified. Generally, KNN contains three steps: (1) computing the distance between the target point and each point in the training dataset; (2) according to the K value that the researcher set previously, picking K data points that remain with the lowest distance with the target; (3) applying a majority vote that the prediction result is in accordance with the majority of class among K neighbor points.

Support Vector Machine (SVM). SVM is a machine learning algorithm generally used for data classification. SVM generates a hyperplane or a group of hyperplanes in a high-dimensional space to distinctly classify the data points. There are many hyperplanes that are available to select, but SVM aims to look for a plane that is within the maximum margin which means this plane will keep the maximum distance between data points of both classes. SVM was initially a binary classification approach while, after the model extension, SVM can be used for multi-class classification. One-vs.-All (One-vs.-Rest) and One-vs.-One approaches are two common ways to solve multi-class tasks. Assume an N-class problem, One-vs.-All approach generates N binary SVM classifiers while each one splits one class from the rest [95–97]. The i th ($i \in N$) classifier is trained with the whole training data points of i th class labeled positive, and with all the other classes labeled negative. One-vs.-One approach is developed based on the idea that a pair of distinct classes are selected each time and are trained by a binary SVM classifier and an ensemble of binary SVM classifiers forms One-vs.-One SVM classifier. Assume a problem of N different classes, $\frac{N(N-1)}{2}$ binary SVM classifiers are required for separating every two distinct classes. Compared to One-vs.-All SVM, One-vs.-One SVM is more computationally expensive since more binary SVM classifiers are created. We will implement both One-vs.-All SVM and One-vs.-One SVM for model evaluation [98]. Figure 8 demonstrates the SVM classifier that divides scaffolding activities into erecting, transporting and waiting/idling three classes.

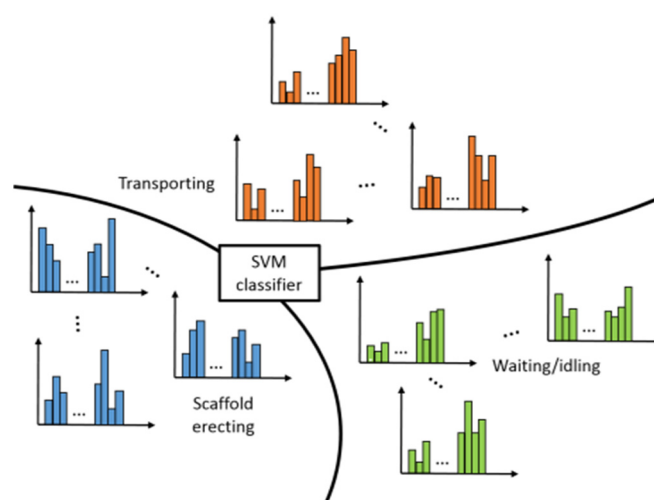


Figure 8. SVM classifier (It looks for a group of hyperplanes to classify data points or vectors).

Artificial Neural Networks (ANN). ANN were inspired by the mechanism and structure of biological neural networks in a human's brain in that the signals are received, transmitted and emitted back and forth through multiple layers of neurons. The basic unit in ANN is called a neuron or node. Each neuron is responsible for receiving input data, merging the input with its own parameters or processing the input with activate function. Activate function plays the role of value conversion that avoiding great output variation as the input changes. Connections link every neuron with weights. A weight reflects the contribution percentage of an input to the output of the next layer neuron. Typically, ANN is initially set up with randomized weights of all the neurons. Generally, neurons

are organized into numerous layers and every neuron in the layer only connects with the neurons in the closest neighboring layers. ANN consists of an input layer where input is received, a hidden layer, and an output layer where the result is produced. The hidden layer is sandwiched between the input and output layers and can be multiple according to the complexity of ANN architecture. At the training stage, ANN adopts the algorithms of forward propagation and backward propagation to train the model. Forward propagation refers to the procedure that the input data is fed in the forward direction from the input through the hidden layers to the output layer in the neural networks and the calculations and the storage of inter parameters are conducted. Backward propagation starts from the comparison of deviation between the expected (labeled) output and the generated output and it runs iteratively in order to minimize this deviation to a predetermined level. As all the weights of neurons are randomly initialized, backward propagation regulates the contribution of every node to the final output by nudging the weight connecting every neuron of layers from the output layer to the input layer. As a result, the NN can predict the desirable output automatically.

For each supervised learning classifier, we optimize the parameters of models and functions by iteratively comparing internal model performance on the same training dataset. The optimal classifier is chosen by picking the supervised classifier with the highest mean accuracy through 10-folds cross-validation, which represents a globally recognized evaluation method of classification performance, and the presentation of confusion matrix with reasonable results. Next, the optimal classifier is designed to conduct prediction of workplace assessment from several video sequences comprising various scaffolding activities in actual working conditions. The predicted workplace assessment is compared with the manual annotation on the same video sequence for validation and discussion.

4. Case Study and Results Analysis

Due to the lack of relevant databases for training and testing visual activities of scaffolding construction operations, before validating our approach it is crucial to collect a corresponding video sequence dataset. As shown in Figures 9 and 10, the target scaffolding operation comprises 3 activities: scaffold erecting, transporting and idling/waiting. Cameras with 12 megapixels were used to capture video data in a real construction environment in Western Australia, as well as in the laboratory of the Australasian Joint Research Centre for Building Information Modelling in Curtin University. Professional volunteers were recruited to conduct the same activities both under the actual and laboratory construction condition. Workplace assessment requires a supervisor to clearly observe and record entire scaffolding activities from a certain distance. This observing distance ranges from 5 m to 20 m, which allows the supervisor to identify distinct activities. Thus, in our approach, portable video cameras were placed 5 m to 20 m away from the scaffolding activities. To increase the variety of the database, the camera shot angle was set at three height levels: the first one was located between ground level and kneel level, the second one was located between kneel level and eye level and the third one was located over the head level. Horizontal camera angles ranged from a frontal angle, a three-quarter front angle to a profile angle (a view from the side). Some 52 clips of video regarding relevant scaffolding activities were collected and some of these videos are trimmed, in that they last more than 8 s and only include one single activity for training and testing purposes, and the rest of the videos covering several scaffolding activities are stored for validation. Some 40 clips were collected in the lab and the remaining 12 clips were collected on-site. The duration of the clips lasts from 3 to 10 min. The captured video has 25 fps and 1080p resolution. Every video clip is processed into a stack of image frames and the whole dataset consists of 1731 frames. Additionally, these frames are annotated with their corresponding activity category including erecting, transporting and idling/waiting as the input for the purpose of supervised learning. Every video clip that merely includes one category of scaffolding activity was processed into image frames. For each image frame, its corresponding activity category was added as one additional feature among 42 features of body key points.

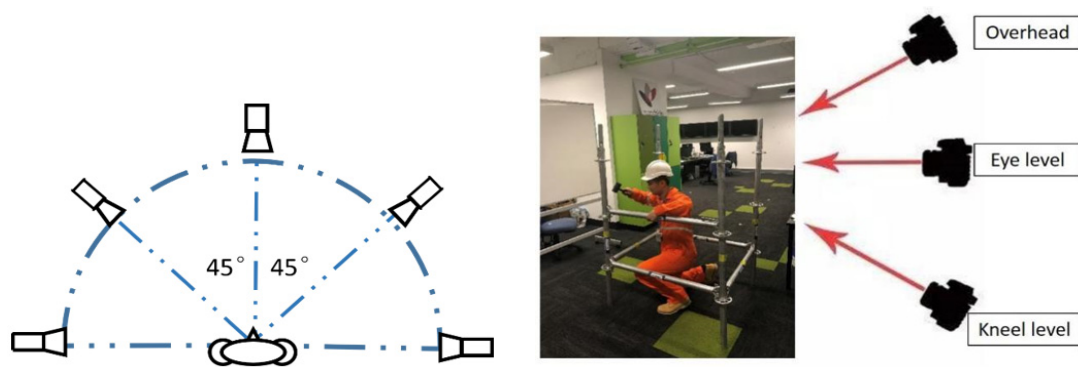


Figure 9. Experiment settings: the camera shot angle was set at three height levels: kneel level, eye level and overhead level and the horizontal camera angles range from the frontal angle, and three-quarter front angle to profile angle.



Figure 10. Video sequences of scaffolding activity: (a) scaffolding erecting; (b) transporting; (c) waiting or idling.

To evaluate the performance of our proposed approach for construction activity analysis, a case study was carried out. The methodology for activity classification was achieved in a Python-based environment and open-source package Scikit-learn with the assistance of a DELL workstation with a 1.9 GHz processing unit and 32 GB RAM.

As it is shown in Table 2, the mean accuracy of selected supervised classifiers under 10-folds cross-validation regarding our scaffolding dataset was 96.58% (RF), 94.24% (SVM), 92.08% (DT), 96.13% (KNN), 93.12% (NN) respectively. All the classifiers' general accu-

racy achieved over 90%, while RF and KNN outperformed the rest with an accuracy of around 96%.

Table 2. The performance of the classifiers.

Classifier	Accuracy	Macroaverage Recall	Macroaverage Precision	Average_F1 Score
RF	96.58%	0.9445	0.9405	0.9422
SVM	94.24%	0.9209	0.9349	0.9273
DT	92.08%	0.9040	0.8749	0.8880
KNN	96.13%	0.9492	0.9578	0.9537
NN	93.12%	0.9339	0.9337	0.9337

Since 3D pose classification is a transitional step in our model, every video frame of scaffolding work can be classified as one of three scaffolding activities. The results of 3D pose estimations include the classification results for every video frame as well as the visualization of 3D key joints and pose estimation. The classification results of every video frame were evaluated and validated via 10-fold cross-validation and the performance metrics including accuracy, precision, recall and F1 score indicated that all the classifiers achieved good performances. The confusion matrix of the classifiers are presented in Figure 11.

Random Forest (RF)	Class1: Working	Class2: Transporting	Class3: Idling
Class1: Working	784	11	9
Class2: Transporting	12	751	8
Class3: Idling	7	13	135

Support Vector Machine (SVM)	Class1: Working	Class2: Transporting	Class3: Idling
Class1: Working	763	21	4
Class2: Transporting	47	734	17
Class3: Idling	9	3	132

Decision Tree (DT)	Class1: Working	Class2: Transporting	Class3: Idling
Class1: Working	746	31	12
Class2: Transporting	42	717	9
Class3: Idling	18	26	129

K-Nearest Neighbor (KNN)	Class1: Working	Class2: Transporting	Class3: Idling
Class1: Working	780	22	5
Class2: Transporting	27	748	7
Class3: Idling	7	0	134

Neural Networks (NN)	Class1: Working	Class2: Transporting	Class3: Idling
Class1: Working	755	41	6
Class2: Transporting	63	725	2
Class3: Idling	1	7	130

Figure 11. Confusion matrix of the classifiers. The shaded boxes indicate the majority of detection results.

To test the capability of the formation of workplace assessment, three video clips containing all activity categories were used for validation. As shown in Table 3, Video Clip 1 lasts for 206 s and Clips 2 and 3 last 116 s and 79 s, respectively. In addition, Clip 1 was converted into a sequence of image frames by extracting frames at 2 s intervals while Clips

2 and 3 were converted with 1 s intervals. 1731 frames were used for training and testing and these frames went through 10-fold cross-validation. Some 314 frames were used for the validation of the case study. For the shot angle, Clip 1 was collected at kneel level, while Clips 2 and 3 were collected at chest level and overhead, respectively, to test the ability of generalization of our model. Since an RF model has the highest mean accuracy and relatively high score in Recall, Precision and F1 score through our training procedure, we chose an RF model for the validation of the workplace assessment. The sequences of image frames from three clips were fed into the OpenPose and 3D extraction model and then 3D pose features were generated and transmitted through our RF classifier for activity classification. Figure 12 shows an example of 3D key joints and skeleton extraction. Each frame was annotated with one of our three activity categories by the RF classifier, and we put these annotations in order of temporal dimension to form the workplace assessment. During the formation, we eliminated the noise of activity status by taking the majority status among each image frame. The duration of each activity category is determined by this logic; that we set 5 s as the unit of time period and picked the majority category (among the categories before and after 2 s) as the activity category that this time point belongs to. Thus, the boundary of each activity category occurs when two different activity categories are identified before and after one certain point in time. For example, one central frame was annotated by the model as the waiting/idling category, while the rest frames next to it were annotated as scaffolding erecting status and we took the erecting as the activity category of this period.

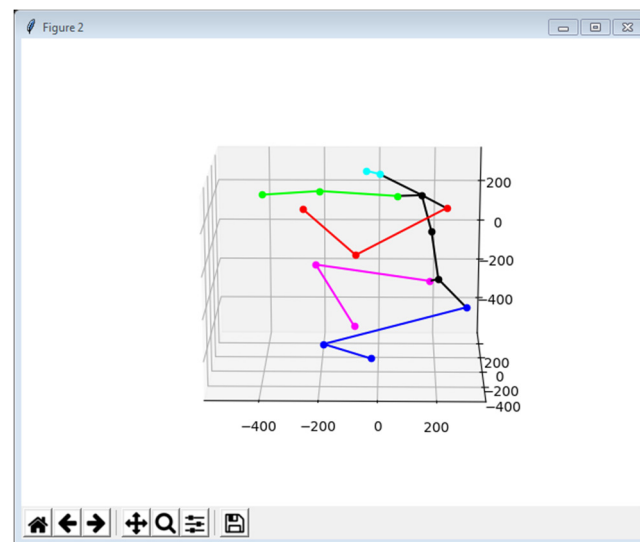
Table 3. Details of video clips for validation.

#Clip	Length (s)	Number of Frames	Frame Interval (s)	Shot Angle
1	03:26	102	2	Kneel level
2	01:56	118	1	Chest level
3	01:19	79	1	Overhead



(a)

Figure 12. Cont.



(b)

Figure 12. 3D key joints and skeleton extraction: (a) key joint distribution in the image frame of scaffolding erecting; (b) the corresponding skeleton model in 3D.

Based on the comparison between the model outputs and the manual observation in Figure 13, this case study displayed several phenomena. Firstly, our model presented the feasibility and capability of discrimination between the scaffolding activities of erecting and transporting, and generated relatively effective intervals between these two activities. Although there exists a little divergence between the boundary of distinct activities, which may be caused by the transitional procedure during the conduct of scaffolding activities or the occlusion of body parts, the model classification is capable of basically reflecting the activity status.

Secondly, the discrimination between the activities erecting and waiting/idling is vulnerable when the participant performed either erecting or waiting/idling activities. In three testing clips, the waiting/idling activity was misclassified as scaffolding erecting. This poor discrimination may be a result of the complexity of the activity of scaffolding erecting or the similarity between these two activities.

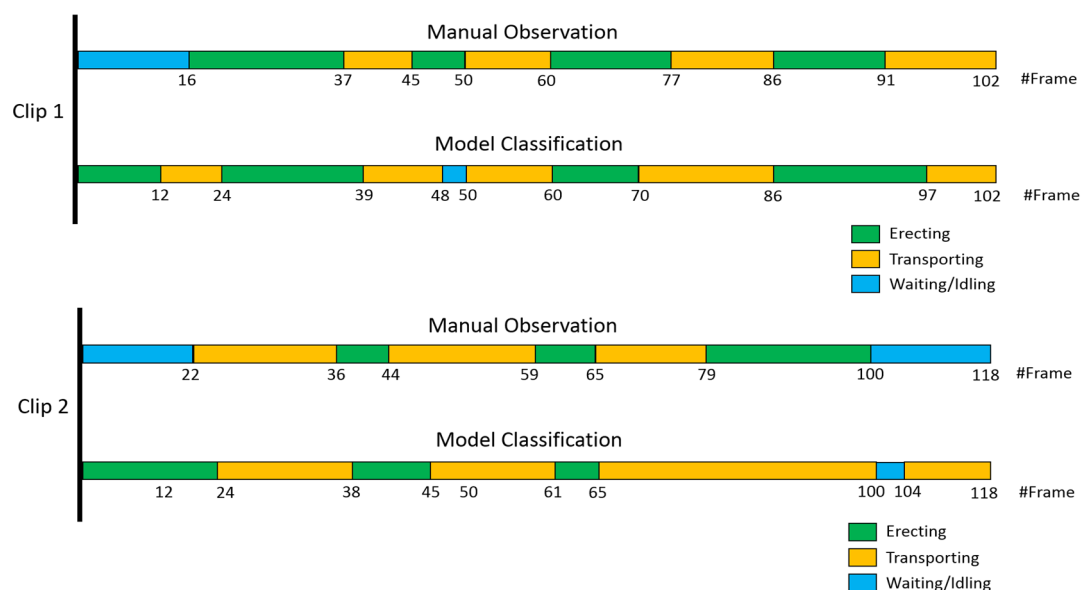


Figure 13. Cont.

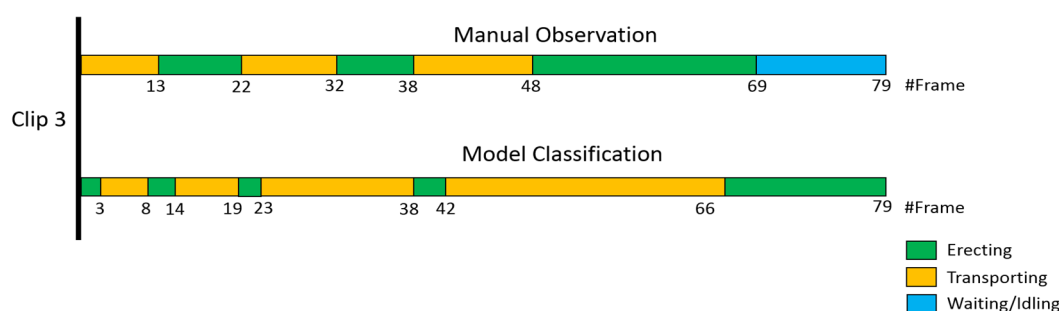


Figure 13. Comparison of workforce assessment between manually labeled ground truth and automatically generated results from the model.

5. Discussion and Conclusions

In conclusion, under the background of a rising interest in automation and digitalization in the construction industry, a method of automatic scaffolding workforce assessment is proposed and developed in this paper to replace repetitive man-hours for data collection and supervision. Scaffolding activities can be recorded and visually monitored throughout onsite video cameras and then classified into three activity categories based on the procedure of workforce assessment. One video dataset of scaffolding activities was created for training and testing. The case study demonstrates the feasibility of our developed model in realistic conditions. The developed model presents enormous potential to enhance project monitoring and control.

At the current stage, the proposed model merely takes every video frame as the input without taking the temporal feature into consideration. As a result, our model classifies every video frame of scaffolding into an individual activity category without considering scaffolding postures in one kind of activity as an entity. We simplified and split scaffolding activity into three categories, however, scaffolding has great complexity in activity transition as well as activity subdivision.

The core contribution of this paper includes: (1) a video dataset is established for the research of scaffolding activities; (2) an approach is developed for automatic workforce assessment of scaffolding by using the method of 3D key joint extraction and machine learning classification; (3) we followed the principle of activity analysis and simplified scaffolding activities into three activities, erecting, transporting and idling, and then explored the potential of activity recognition for the field of construction inspection and productivity improvement. This approach is designed to release on-site managers from regular inspections with the help of video surveillance.

Future work would be placed on increasing the volume and variety of training data to increase the robustness of model prediction. Meanwhile, improving the structure of our prediction model is another area that requires great efforts in the future. In addition, due to the complex nature of the scaffolding activity itself, the scaffolding activity needs to be analyzed and divided indepth, which would also be one important direction in our future research.

Author Contributions: Conceptualization, W.Y., W.S. (Wenchi Shou), J.W. and X.W.; funding acquisition, X.W.; methodology, W.Y., and H.G.; project administration, M.C.; software, W.Y. and Y.S.; supervision, X.W. and M.C.; Visualization, W.Y. and D.J.; writing—original draft, W.Y.; writing—review & editing, W.Y., W.S. (Wenchi Shou) and W.S. (Weixiang Shi). All authors have read and agreed to the published version of the manuscript.

Funding: This research was jointly funded by ARC DP 170104612 and ARC DP 170104613 from Australian Research Council, Australia, and Industry Doctoral Training Centre (IDTC) from Australian Technology Network of Universities (ATN).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Human Research Ethics Office, Curtin University (Approval number: HRE2019-0797; Date: 22-11-2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hu, X.; Chong, H.-Y.; Wang, X. Sustainability perceptions of off-site manufacturing stakeholders in Australia. *J. Clean. Prod.* **2019**, *227*, 346–354. [\[CrossRef\]](#)
2. Ye, M.; Jiang, J.; Chen, H.; Zhou, H.; Song, D. Seismic behavior of an innovative hybrid beam-column connection for precast concrete structures. *Eng. Struct.* **2021**, *227*, 111436. [\[CrossRef\]](#)
3. Huang, H.; Guo, M.; Zhang, W.; Zeng, J.; Yang, K.; Bai, H. Numerical investigation on the bearing capacity of RC columns strengthened by HPFL-BSP under combined loadings. *J. Build. Eng.* **2021**, *39*, 102266. [\[CrossRef\]](#)
4. Huang, H.; Huang, M.; Zhang, W.; Yang, S. Experimental study of predamaged columns strengthened by HPFL and BSP under combined load cases. *Struct. Infrastruct. Eng.* **2020**, 1–18. [\[CrossRef\]](#)
5. Moon, S.; Forlani, J.; Wang, X.; Tam, V. Productivity Study of the Scaffolding Operations in Liquefied Natural Gas Plant Construction: Ichthys Project in Darwin, Northern Territory, Australia. *J. Prof. Issues Eng. Educ. Pract.* **2016**, *142*, 04016008. [\[CrossRef\]](#)
6. Chen, Y.; Zheng, W.; Li, W.; Huang, Y. Large group activity security risk assessment and risk early warning based on random forest algorithm. *Pattern Recognit. Lett.* **2021**, *144*, 1–5. [\[CrossRef\]](#)
7. Hou, L.; Zhao, C.; Wu, C.; Moon, S.; Wang, X. Discrete Firefly Algorithm for Scaffolding Construction Scheduling. *J. Comput. Civ. Eng.* **2017**, *31*, 04016064. [\[CrossRef\]](#)
8. Kim, K.; Teizer, J. Automatic design and planning of scaffolding systems using building information modeling. *Adv. Eng. Inform.* **2014**, *28*, 66–80. [\[CrossRef\]](#)
9. Zhu, J.; Wang, X.; Chen, M.; Wu, P.; Kim, M.J. Integration of BIM and GIS: IFC geometry transformation to shapefile using enhanced open-source approach. *Autom. Constr.* **2019**, *106*, 102859. [\[CrossRef\]](#)
10. Singh, V.; Gu, N.; Wang, X. A theoretical framework of a BIM-based multi-disciplinary collaboration platform. *Autom. Constr.* **2011**, *20*, 134–144. [\[CrossRef\]](#)
11. Chong, H.-Y.; Lee, C.-Y.; Wang, X. A mixed review of the adoption of Building Information Modelling (BIM) for sustainability. *J. Clean. Prod.* **2017**, *142*, 4114–4126. [\[CrossRef\]](#)
12. Zhu, J.; Wu, P.; Chen, M.; Kim, M.J.; Wang, X.; Fang, T. Automatically Processing IFC Clipping Representation for BIM and GIS Integration at the Process Level. *Appl. Sci.* **2020**, *10*, 2009. [\[CrossRef\]](#)
13. Omar, H.; Mahdjoubi, L.; Kheder, G. Towards an automated photogrammetry-based approach for monitoring and controlling construction site activities. *Comput. Ind.* **2018**, *98*, 172–182. [\[CrossRef\]](#)
14. Wang, L.; Yuan, J.; Wu, C.; Wang, X. Practical algorithm for stochastic optimal control problem about microbial fermentation in batch culture. *Optim. Lett.* **2019**, *13*, 527–541. [\[CrossRef\]](#)
15. Golparvar-Fard, M.; Bohn, J.; Teizer, J.; Savarese, S.; Peña-Mora, F. Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques. *Autom. Constr.* **2011**, *20*, 1143–1155. [\[CrossRef\]](#)
16. Bosché, F.; Guillemet, A.; Turkan, Y.; Haas, C.T.; Haas, R. Tracking the Built Status of MEP Works: Assessing the Value of a Scan-vs-BIM System. *J. Comput. Civ. Eng.* **2014**, *28*, 05014004. [\[CrossRef\]](#)
17. Wu, P.; Wang, J.; Wang, X. A critical review of the use of 3-D printing in the construction industry. *Autom. Constr.* **2016**, *68*, 21–31. [\[CrossRef\]](#)
18. Li, X.; Yi, W.; Chi, H.-L.; Wang, X.; Chan, A.P. A critical review of virtual and augmented reality (VR/AR) applications in construction safety. *Autom. Constr.* **2018**, *86*, 150–162. [\[CrossRef\]](#)
19. Niu, Z.; Zhang, B.; Wang, J.; Liu, K.; Chen, Z.; Yang, K.; Zhou, Z.; Fan, Y.; Zhang, Y.; Ji, D.; et al. The research on 220GHz multicarrier high-speed communication system. *China Commun.* **2020**, *17*, 131–139. [\[CrossRef\]](#)
20. Gouett, M.C.; Haas, C.T.; Goodrum, P.M.; Caldas, C.H. Activity Analysis for Direct-Work Rate Improvement in Construction. *J. Constr. Eng. Manag.* **2011**, *137*, 1117–1124. [\[CrossRef\]](#)
21. CII. *Guide to Activity Analysis*; University of Texas at Austin: Austin, TX, USA, 2010.
22. Khosrowpour, A.; Niebles, J.C.; Golparvar-Fard, M. Vision-based workplace assessment using depth images for activity analysis of interior construction operations. *Autom. Constr.* **2014**, *48*, 74–87. [\[CrossRef\]](#)
23. Ma, X.; Zhang, K.; Zhang, L.; Yao, C.; Yao, J.; Wang, H.; Jian, W.; Yan, Y. Data-Driven Niching Differential Evolution with Adaptive Parameters Control for History Matching and Uncertainty Quantification. *SPE J.* **2021**, *26*, 993–1010. [\[CrossRef\]](#)
24. Shi, J.; Lu, Y.; Zhang, J. Approximation Attacks on Strong PUFs. *IEEE Trans. Comput. Des. Integr. Circuits Syst.* **2020**, *39*, 2138–2151. [\[CrossRef\]](#)

25. Zhang, J.; Shen, C.; Su, H.; Arafat, T.; Qu, G. Voltage Over-scaling-based Lightweight Authentication for IoT Security. *IEEE Trans. Comput.* **2021**, *1*. [\[CrossRef\]](#)
26. Cheng, T.; Teizer, J.; Migliaccio, G.C.; Gatti, U.C. Automated task-level activity analysis through fusion of real time location sensors and worker's thoracic posture data. *Autom. Constr.* **2013**, *29*, 24–39. [\[CrossRef\]](#)
27. Zhang, B.; Ji, D.; Fang, D.; Liang, S.; Fan, Y.; Chen, X. A Novel 220-GHz GaN Diode On-Chip Tripler with High Driven Power. *IEEE Electron Device Lett.* **2019**, *40*, 780–783. [\[CrossRef\]](#)
28. Liu, Y.; Zhang, B.; Feng, Y.; Lv, X.; Ji, D.; Niu, Z.; Yang, Y.; Zhao, X.; Fan, Y. Development of 340-GHz Transceiver Front End Based on GaAs Monolithic Integration Technology for THz Active Imaging Array. *Appl. Sci.* **2020**, *10*, 7924. [\[CrossRef\]](#)
29. Yang, Y.; Hou, C.; Lang, Y.; Sakamoto, T.; He, Y.; Xiang, W. Omnidirectional Motion Classification with Monostatic Radar System Using Micro-Doppler Signatures. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 3574–3587. [\[CrossRef\]](#)
30. Jiang, Q.; Shao, F.; Gao, W.; Chen, Z.; Jiang, G.; Ho, Y.-S. Unified No-Reference Quality Assessment of Singly and Multiply Distorted Stereoscopic Images. *IEEE Trans. Image Process.* **2019**, *28*, 1866–1881. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Zhang, J.; Chen, Q.; Sun, J.; Tian, L.; Zuo, C. On a universal solution to the transport-of-intensity equation. *Opt. Lett.* **2020**, *45*, 3649. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Zhang, J.; Sun, J.; Chen, Q.; Zuo, C. Resolution Analysis in a Lens-Free On-Chip Digital Holographic Microscope. *IEEE Trans. Comput. Imaging* **2020**, *6*, 697–710. [\[CrossRef\]](#)
33. Assadzadeh, A.; Arashpour, M.; Bab-Hadiashar, A.; Ngo, T.; Li, H. Automatic far-field camera calibration for construction scene analysis. *Comput. Civ. Infrastruct. Eng.* **2021**. [\[CrossRef\]](#)
34. Jiang, Q.; Shao, F.; Lin, W.; Gu, K.; Jiang, G.; Sun, H. Optimizing Multistage Discriminative Dictionaries for Blind Image Quality Assessment. *IEEE Trans. Multimed.* **2018**, *20*, 2035–2048. [\[CrossRef\]](#)
35. Li, B.; Liang, R.; Zhou, W.; Yin, H.; Gao, H.; Cai, K. LBS Meets Blockchain: an Efficient Method with Security Preserving Trust in SAGIN. *IEEE Internet Things J.* **2021**, *1*. [\[CrossRef\]](#)
36. Sarafianos, N.; Boteanu, B.; Ionescu, B.; Kakadiaris, I.A. 3D Human pose estimation: A review of the literature and analysis of covariates. *Comput. Vis. Image Underst.* **2016**, *152*, 1–20. [\[CrossRef\]](#)
37. Dang, L.M.; Min, K.; Wang, H.; Piran, J.; Lee, C.H.; Moon, H. Sensor-based and vision-based human activity recognition: A comprehensive survey. *Pattern Recognit.* **2020**, *108*, 107561. [\[CrossRef\]](#)
38. Niu, Z.-Q.; Zhang, B.; Li, D.-T.; Ji, D.-F.; Liu, Y.; Feng, Y.-N.; Zhou, T.-C.; Zhang, Y.-H.; Fan, Y. A mechanical reliability study of 3dB waveguide hybrid couplers in the submillimeter and terahertz band. *J. Zhejiang Univ. Sci.* **1998**, *1*, 1. [\[CrossRef\]](#)
39. Ryu, J.; Seo, J.; Jebelli, H.; Lee, S. Automated Action Recognition Using an Accelerometer-Embedded Wristband-Type Activity Tracker. *J. Constr. Eng. Manag.* **2019**, *145*, 04018114. [\[CrossRef\]](#)
40. Bangaru, S.S.; Wang, C.; Aghazadeh, F. Data Quality and Reliability Assessment of Wearable EMG and IMU Sensor for Construction Activity Recognition. *Sensors* **2020**, *20*, 5264. [\[CrossRef\]](#)
41. Bangaru, S.S.; Wang, C.; Busam, S.A.; Aghazadeh, F. ANN-based automated scaffold builder activity recognition through wearable EMG and IMU sensors. *Autom. Constr.* **2021**, *126*, 103653. [\[CrossRef\]](#)
42. Zuo, C.; Chen, Q.; Tian, L.; Waller, L.; Asundi, A. Transport of intensity phase retrieval and computational imaging for partially coherent fields: The phase space perspective. *Opt. Lasers Eng.* **2015**, *71*, 20–32. [\[CrossRef\]](#)
43. Fiğlalı, N.; Cihan, A.; Esen, H.; Fiğlalı, A.; Çeşmeci, D.; Güllü, M.K.; Yılmaz, M.K.; Yılmaz, M.K. Image processing-aided working posture analysis: I-OWAS. *Comput. Ind. Eng.* **2015**, *85*, 384–394. [\[CrossRef\]](#)
44. Rahman, C. Study and analysis of work postures of workers working in a ceramic industry through rapid upper limb assessment (RULA). *Int. J. Eng.* **2014**, *5*, 8269.
45. Ray, S.J.; Teizer, J. Real-time construction worker posture analysis for ergonomics training. *Adv. Eng. Inform.* **2012**, *26*, 439–455. [\[CrossRef\]](#)
46. Yan, X.; Li, H.; Wang, C.; Seo, J.; Zhang, H.; Wang, H. Development of ergonomic posture recognition technique based on 2D ordinary camera for construction hazard prevention through view-invariant features in 2D skeleton motion. *Adv. Eng. Inform.* **2017**, *34*, 152–163. [\[CrossRef\]](#)
47. Zhang, H.; Yan, X.; Li, H. Ergonomic posture recognition using 3D view-invariant features from single ordinary camera. *Autom. Constr.* **2018**, *94*, 1–10. [\[CrossRef\]](#)
48. Wang, P.; Wu, P.; Wang, X.; Chen, X.; Zhou, T. Developing optimal scaffolding erection through the integration of lean and work posture analysis. *Eng. Constr. Arch. Manag.* **2020**, *27*, 2109–2133. [\[CrossRef\]](#)
49. Yan, X.; Li, H.; Li, A.R.; Zhang, H. Wearable IMU-based real-time motion warning system for construction workers' musculoskeletal disorders prevention. *Autom. Constr.* **2017**, *74*, 2–11. [\[CrossRef\]](#)
50. Seo, J.; Starbuck, R.; Han, S.; Lee, S.; Armstrong, T.J. Motion Data-Driven Biomechanical Analysis during Construction Tasks on Sites. *J. Comput. Civ. Eng.* **2015**, *29*, 4014005. [\[CrossRef\]](#)
51. Yu, Y.; Guo, H.; Ding, Q.; Li, H.; Skitmore, M. An experimental study of real-time identification of construction workers' unsafe behaviors. *Autom. Constr.* **2017**, *82*, 193–206. [\[CrossRef\]](#)
52. Zhao, J.; Liu, J.; Jiang, J.; Gao, F. Efficient Deployment with Geometric Analysis for mmWave UAV Communications. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 1. [\[CrossRef\]](#)
53. Alwasel, A.; Sabet, A.; Nahangi, M.; Haas, C.T.; Abdel-Rahman, E. Identifying poses of safe and productive masons using machine learning. *Autom. Constr.* **2017**, *84*, 345–355. [\[CrossRef\]](#)

54. Han, S.; Lee, S.; Peña-Mora, F. Vision-Based Detection of Unsafe Actions of a Construction Worker: Case Study of Ladder Climbing. *J. Comput. Civ. Eng.* **2013**, *27*, 635–644. [\[CrossRef\]](#)
55. Zhang, B.; Niu, Z.; Wang, J.; Ji, D.; Zhou, T.; Liu, Y.; Feng, Y.; Hu, Y.; Zhang, J.; Fan, Y. Four-hundred gigahertz broadband multi-branch waveguide coupler. *IET Microw. Antennas Propag.* **2020**, *14*, 1175–1179. [\[CrossRef\]](#)
56. Luo, H.; Xiong, C.; Fang, W.; Love, P.E.; Zhang, B.; Ouyang, X. Convolutional neural networks: Computer vision-based workforce activity assessment in construction. *Autom. Constr.* **2018**, *94*, 282–289. [\[CrossRef\]](#)
57. Hu, J.; Zhang, H.; Liu, L.; Zhu, X.; Zhao, C.; Pan, Q. Convergent Multiagent Formation Control with Collision Avoidance. *IEEE Trans. Robot.* **2020**, *36*, 1805–1818. [\[CrossRef\]](#)
58. Peddi, A.; Huan, L.; Bai, Y.; Kim, S. Development of Human Pose Analyzing Algorithms for the Determination of Construction Productivity in Real-Time. In Proceedings of the Construction Research Congress, Seattle, WA, USA, 5–7 April 2009.
59. Calderon, W.T.; Roberts, D.; Golparvar-Fard, M. Synthesizing Pose Sequences from 3D Assets for Vision-Based Activity Analysis. *J. Comput. Civ. Eng.* **2021**, *35*, 04020052. [\[CrossRef\]](#)
60. Guraliuc, A.R.; Barsocchi, P.; Potorti, F.; Nepa, P. Limb Movements Classification Using Wearable Wireless Transceivers. *IEEE Trans. Inf. Technol. Biomed.* **2011**, *15*, 474–480. [\[CrossRef\]](#)
61. Amendola, S.; Bianchi, L.; Marrocco, G. Movement Detection of Human Body Segments: Passive radio-frequency identification and machine-learning technologies. *IEEE Antennas Propag. Mag.* **2015**, *57*, 23–37. [\[CrossRef\]](#)
62. Joshua, L.; Varghese, K. Accelerometer-Based Activity Recognition in Construction. *J. Comput. Civ. Eng.* **2011**, *25*, 370–379. [\[CrossRef\]](#)
63. Aloui, S.; Villien, C.; Lesecq, S. A new approach for motion capture using magnetic field: Models, algorithms and first results. *Int. J. Adapt. Control. Signal Process.* **2014**, *29*, 407–426. [\[CrossRef\]](#)
64. Pradhananga, N.; Teizer, J. Automatic spatio-temporal analysis of construction site equipment operations using GPS data. *Autom. Constr.* **2013**, *29*, 107–122. [\[CrossRef\]](#)
65. Costin, A.; Pradhananga, N.; Teizer, J. Leveraging passive RFID technology for construction resource field mobility and status monitoring in a high-rise renovation project. *Autom. Constr.* **2012**, *24*, 1–15. [\[CrossRef\]](#)
66. Cheng, T.; Venugopal, M.; Teizer, J.; Vela, P.A. Performance evaluation of ultra wideband technology for construction resource location tracking in harsh environments. *Autom. Constr.* **2011**, *20*, 1173–1184. [\[CrossRef\]](#)
67. Hildreth, J.; Vorster, M.; Martinez, J. Reduction of Short-Interval GPS Data for Construction Operations Analysis. *J. Constr. Eng. Manag.* **2005**, *131*, 920–927. [\[CrossRef\]](#)
68. Teizer, J.; Vela, P. Personnel tracking on construction sites using video cameras. *Adv. Eng. Inform.* **2009**, *23*, 452–462. [\[CrossRef\]](#)
69. Azar, E.R.; McCabe, B. Vision-Based Recognition of Dirt Loading Cycles in Construction Sites. In Proceedings of the Construction Research Congress, West Lafayette, IN, USA, 21–23 May 2012; pp. 1042–1051. [\[CrossRef\]](#)
70. Hu, Y.; Chen, Q.; Feng, S.; Zuo, C. Microscopic fringe projection profilometry: A review. *Opt. Lasers Eng.* **2020**, *135*, 106192. [\[CrossRef\]](#)
71. Wu, C.; Wang, X.; Chen, M.; Kim, M.J. Differential received signal strength based RFID positioning for construction equipment tracking. *Adv. Eng. Inform.* **2019**, *42*, 100960. [\[CrossRef\]](#)
72. Zhou, Y.; Tian, L.; Zhu, C.; Jin, X.; Sun, Y. Video Coding Optimization for Virtual Reality 360-Degree Source. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 118–129. [\[CrossRef\]](#)
73. Ali, S.; Shah, M. Human Action Recognition in Videos Using Kinematic Features and Multiple Instance Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *32*, 288–303. [\[CrossRef\]](#)
74. Sheikh, Y.; Sheikh, M.; Shah, M. Exploring the space of a human action. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, 17–21 October 2005.
75. Jiang, H.; Drew, M.S.; Li, Z.-N. Successive Convex Matching for Action Detection. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), New York, NY, USA, 17–22 June 2006.
76. Liu, J.; Ali, S.; Shah, M. Recognizing human actions using multiple features. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
77. Zuo, C.; Sun, J.; Jialin, Z.; Zhang, J.; Asundi, A.; Chen, Q. High-resolution transport-of-intensity quantitative phase microscopy with annular illumination. *Sci. Rep.* **2017**, *7*, 1–22. [\[CrossRef\]](#) [\[PubMed\]](#)
78. Khosrowpour, A.; Fedorov, I.; Holynski, A.; Niebles, J.C.; Golparvar-Fard, M. Automated Worker Activity Analysis in Indoor Environments for Direct-Work Rate Improvement from Long Sequences of RGB-D Images. In Proceedings of the Construction Research Congress, Atlanta, GA, USA, 19–21 May 2014; pp. 729–738. [\[CrossRef\]](#)
79. Thomas, H.R.; Daily, J. Crew Performance Measurement Via Activity Sampling. *J. Constr. Eng. Manag.* **1983**, *109*, 309–320. [\[CrossRef\]](#)
80. Shou, W.; Wang, J.; Wu, P.; Wang, X.; Chong, H.-Y. A cross-sector review on the use of value stream mapping. *Int. J. Prod. Res.* **2017**, *55*, 3906–3928. [\[CrossRef\]](#)
81. Cao, Z.; Hidalgo Martinez, G.; Simon, T.; Wei, S.-E.; Sheikh, Y.A. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *1*, 172–186. [\[CrossRef\]](#) [\[PubMed\]](#)
82. Zhang, J.; Shen, C. Set-Based Obfuscation for Strong PUFs Against Machine Learning Attacks. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2021**, *68*, 288–300. [\[CrossRef\]](#)
83. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.

84. Yang, S.; Gao, T.; Wang, J.; Deng, B.; Lansdell, B.; Linares-Barranco, B. Efficient Spike-Driven Learning with Dendritic Event-Based Processing. *Front. Neurosci.* **2021**, *15*, 97. [\[CrossRef\]](#)
85. Ionescu, C.; Papava, D.; Olaru, V.; Sminchisescu, C. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1325–1339. [\[CrossRef\]](#)
86. Xue, X.; Zhang, K.; Tan, K.C.; Feng, L.; Wang, J.; Chen, G.; Zhao, X.; Zhang, L.; Yao, J. Affine Transformation-Enhanced Multifactorial Optimization for Heterogeneous Problems. *IEEE Trans. Cybern.* **2020**, 1–15. [\[CrossRef\]](#)
87. Pitelis, N.; Russell, C.; Agapito, L. Learning a manifold as an atlas. In Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 1325–1339.
88. Tome, D.; Russell, C.; Agapito, L. Lifting from the Deep: Convolutional 3D Pose Estimation from a Single Image. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5689–5698.
89. Zhang, J.; Qu, G. Physical Unclonable Function-Based Key Sharing via Machine Learning for IoT Security. *IEEE Trans. Ind. Electron.* **2019**, *67*, 7025–7033. [\[CrossRef\]](#)
90. Zhang, K.; Zhang, J.; Ma, X.; Yao, C.; Zhang, L.; Yang, Y.; Wang, J.; Yao, J.; Zhao, H. History Matching of Naturally Fractured Reservoirs Using a Deep Sparse Autoencoder. *SPE J.* **2021**, 1–21. [\[CrossRef\]](#)
91. Ding, L.; Huang, L.; Li, S.; Gao, H.; Deng, H.; Li, Y.; Liu, G. Definition and Application of Variable Resistance Coefficient for Wheeled Mobile Robots on Deformable Terrain. *IEEE Trans. Robot.* **2020**, *36*, 894–909. [\[CrossRef\]](#)
92. Shaikhina, T.; Lowe, D.; Daga, S.; Briggs, D.; Higgins, R.; Khovanova, N. Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation. *Biomed. Signal Process. Control.* **2019**, *52*, 456–462. [\[CrossRef\]](#)
93. Li, B.-H.; Liu, Y.; Zhang, A.-M.; Wang, W.-H.; Wan, S. A Survey on Blocking Technology of Entity Resolution. *J. Comput. Sci. Technol.* **2020**, *35*, 769–793. [\[CrossRef\]](#)
94. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [\[CrossRef\]](#)
95. Li, H.; Chen, D.; Zhang, H.; Wu, C.; Wang, X. Hamiltonian analysis of a hydro-energy generation system in the transient of sudden load increasing. *Appl. Energy* **2017**, *185*, 244–253. [\[CrossRef\]](#)
96. Qu, X.; Yu, Y.; Zhou, M.; Lin, C.-T.; Wang, X. Jointly dampening traffic oscillations and improving energy consumption with electric, connected and automated vehicles: A reinforcement learning based approach. *Appl. Energy* **2020**, *257*, 114030. [\[CrossRef\]](#)
97. Song, Y.; Wang, X.; Wright, G.; Thatcher, D.; Wu, P.; Felix, P. Traffic Volume Prediction with Segment-Based Regression Kriging and its Implementation in Assessing the Impact of Heavy Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 232–243. [\[CrossRef\]](#)
98. Hsu, C.-W.; Lin, C.-J. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Networks* **2002**, *13*, 415–425. [\[CrossRef\]](#)